



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Report for Detection of Biothreat Agents and Environmental Samples using the LLNL Virulence Array for DHS

C. Jaing, S. Gardner, K. McLoughlin, J. Thissen,
P. Jackson

April 21, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

**Report for Detection of Biothreat Agents and Environmental Samples using the LLNL
Virulence Array for DHS**

Contributors:

Crystal Jaing
Shea Gardner
Kevin McLoughlin
James Thissen

Lawrence Livermore National Laboratory (LLNL), Livermore, CA

Principal Investigator and Correspondent

Crystal Jaing
925-424-6574, jaing2@llnl.gov

Paul Jackson
(925) 424-2725, jackson80@llnl.gov

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Introduction

The objective of this project is to provide DHS a comprehensive evaluation of the current genomic technologies including genotyping, Taqman PCR, multiple locus variable tandem repeat analysis (MLVA), microarray and high-throughput DNA sequencing in the analysis of biothreat agents from complex environmental samples. This report focuses on the design, testing and results of samples on the Virulence Array.

The original Virulence Array developed was funded by DHS in 2006, and updated to include more biothreat agents with funding from NBACC in 2008. This array contains probes from virulence and antibiotic resistance genes in eight bacterial and eight viral agents, probes for forensic level discrimination for these agents, vector probes to detect evidence of genetic engineering and probes that detect sequences from current BioWatch/LRN PCR amplicons. The array has been previously tested using a variety of strains and near neighbors from these bacterial and viral agents. The data were analyzed using our novel maximum likelihood software. For most of the organisms tested, we have achieved at least species level discrimination.

In this project, we used this array to test various strains from *B. anthracis*, *B. thuringiensis*, BioWatch aerosol filter extracts or soil samples that were spiked with *B. anthracis*, and samples that were previously collected during DHS and EPA environmental release exercises that were known to contain *B. thuringiensis* spores. The array was very sensitive. We were able to detect signatures specific to *B. anthracis* Ames at the 10-100 genome copy level after spiking *B. anthracis* Ames DNA into BioWatch aerosol filter extracts or soil samples and performing whole genome amplification.

METHODS

1. Sequence analysis and microarray probe design to develop "mechanism" chip

1.1 Probe design for virulence and antibiotic resistance gene families (HMM)

We selected target sequences from the genomes by searching for virulence-related proteins using 712 sets of profile hidden Markov models (HMMs). HMM sets were designed or selected by Swan *et al.* to recognize a collection of several hundred virulence-associated protein families identified from the literature and public databases. There are a total of 574 virulence families present in the 8 bacterial and 8 viral agents (Table 1), totaling 41,535 gene sequences. We selected probes so that each target gene sequence would be covered by at least 13 probes, favoring probes that were conserved among the sequences within that gene family. Our original HMMs did not represent genes in VEE or West Nile viruses. For these, we downloaded the 27 profile HMMs in the PFAM database for Togaviridae (VEE) and Flaviviridae (West Nile virus) and searched all available complete and partial Togaviridae and Flaviviridae sequences, respectively, resulting in an additional 34,082 gene sequences, for which we designed probes that provided coverage of at least 13 probes per target sequence. The algorithms used for probe design were described as in (1).

Table 1. Bacterial and viral agents included on the “Virulence Array”

Bacterial Agents	Viral Agents
<i>Bacillus anthracis</i>	Marburg virus
<i>Yersinia pestis</i>	Ebola virus (Reston, Zaire, Sudan)
<i>Francisella tularensis</i>	Variola virus
<i>Burkholderia mallei</i>	Foot-and-Mouth Disease (FMDV) virus
<i>Burkholderia pseudomallei</i>	Venezuelan Equine Encephalitis (VEEV) virus
<i>Brucella abortus</i>	Crimean Congo Haemorrhagic Fever (CCHF) virus
<i>Brucella melintensis</i>	West Nile virus
<i>Brucella suis</i>	Rift Valley Fever (RVF) virus

1.2 Probe design for forensic level discrimination of bacterial and viral strains and *Francisella tularensis*/ *Francisella philomirigia* discriminating genes

We designed microarray probes that span the PCR amplicons from the forensic signatures we recently developed for DHS/NBFAC for species or strain-level discrimination. These signatures were designed using forensic Minimal Set Clustering (MSC) methodology that finds multiple signatures to discriminate all known (sequenced) isolates of a particular agent. This approach determines the maximal target resolution, possibly to a single sequence. MSC typically finds more target resolution for agents with high sequence variation, such as RNA viruses.

We created a non-redundant gene set of ~7,000 genes from the 8 Genbank reference genomes for *Francisella tularensis* tularensis (FTT)/holarctica (FTH)/novicida (FTN) and *Francisella philomirigia* (FP). We then created gene presence/absence vectors against the 7,000-gene set for each of the reference genomes. We designed microarray probes for the set of ~400 total genes that will provide redundant discrimination between FTT, FTH, FTN, and FP strains. Despite the fact that the virulence factors of *Francisella* remain largely unknown, the whole-gene analysis provides a confident way to characterize unknown isolates.

1.3 Probe design for BW/LRN amplicons

We designed microarray probes that span the PCR amplicons from the current BioWatch and LRN signatures. These probes will serve as a secondary confirmation when there are any near positive BioWatch events.

1.4 Probe design for bacterial vectors

We have developed a database with 3,800 complete and partially sequenced vectors and designed microarray probes from unique regions of each vector sequences, using a target goal of 12 probes per vector. Probes are chosen in decreasing order of conservation across the vector sequence database. The initial candidate probe set is screened *in silico* against all sequenced viral and bacterial genomes including naturally occurring plasmids. Candidate vector probes with a similarity above a fixed threshold were removed from the probe set. A cross validation procedure was used to select the threshold to limit the *in silico* predicted false positive rate to 0 while maintaining a predicted high true positive detection rate of 98%, additional details are given in (2).

Table 2. Types of microarray probes for the Virulence Array

Probe Type	# of Probes
BW genus mset	83,372
BW species genomic	842
Viral mset	40,230
Random controls	2,898
BW/LRN_Amplicon	502
Forensic viral probes	1,677
Forensic bacteria probes	1,260
Vector probes	35,791
FT/FP discriminating genes	21,890
Total # of probes	188,462

2. DNA Extraction from pure bacteria and environmental samples

2.1 DNA extraction from *B. anthracis* and *B. thuringiensis* strains.

Genomic DNA from *B. anthracis* Ames, Sterne and A0382 and *B.thuringiensis* Israrensis HD500, kurstaki ATCC 33679 were obtained from LLNL collections.

2.2 DNA extraction from BioWatch Filters

PSU filters from the NCR Laboratory were received from the BioWatch group at LLNL. One quarter of each filter had previously been excised at NCR, so only $\frac{3}{4}$ of each filter was available. One week of filters was collected from each season: Spring (4/20-4/26/09), Summer (7/19-7/25/09), Fall (10/25-10/31/09), and Winter (1/22/09-1/28/09). For each day, 7-11 “clean” filters were extracted (49-77 per week). Filters were determined to be “dirty” if they had an abundance of soot and dirt captured on their surface.

The $\frac{3}{4}$ PSU filters were cut into 5 roughly equal pieces using sterile equipment. Up to 24 filters were placed into a 50mL conical tube. 30mL of 100mM phosphate buffer (pH 7.4), 0.05% (v/v) Tween 80 was added to each 50mL tube. The conical tubes were vortexed for 30 seconds and placed on a rocking shaker for 15 minutes. The 30 second vortexing and 15 min shaking was repeated an additional 3 times for a total of 1 hour of washing. The filters were removed from the tube and remaining solution was centrifuged at $3200 \times g$ for 30 minutes at 5°C. Following centrifugation, the supernatant was removed and discarded.

To complete the DNA purification, components of the UltraClean Soil DNA Isolation Kit #12800 from MoBio (Carlsbad, CA) were utilized. The remaining pellet was resuspended with the following solutions added in this order: 100µL TE buffer, 350µL MoBio Bead Solution, 60µL MoBio Solution S1, and 200µL MoBio Inhibitor Removal Solution. A 2mL screw cap tube was loaded with 500mg each of 106 and 500mm zirconia/silica beads. The entire 700µL of resuspended pellet was added to the 2mL bead tube. The samples were bead-beated at max speed for 2 minutes. Following bead-beating, the tubes were centrifuged at $10,000 \times g$ for 30 seconds. The entire supernatant (~450µL) was transferred to a sterile 2mL tube for further extraction.

To the supernatant, 250 μ L of MoBio Solution S2 was added, vortexed for 5 seconds, and incubated at 4 $^{\circ}$ C for 5 minutes. Following incubation the samples were centrifuged for 1 minute at 10,000 \times g and the supernatant transferred to a clean 2mL tube. 2 volumes (~1.3mL) of MoBio Solution S3 was added to the supernatant and vortexed for 5 seconds. The vortexed solution was added in 700 μ L aliquots, until the entire sample is processed, to a MoBio spin filter and centrifuged for 1 minute at 10,000 \times g and the flow-through discarded. The spin filter was washed 3 times by adding 300 μ L MoBio Solution S4, centrifuged for 30 seconds at 10,000 \times g, and the flow-through discarded.

The spin filter was centrifuged an additional 1 minute at 10,000 \times g to dry the filter. The filter was placed in a new 2mL collection tube and 50 μ L of MoBio Solution S5 was added to the membrane. The sample was centrifuged at 10,000 \times g for 30 seconds and the eluted DNA was retained. The multiple elutions for each season were combined into one large volume. Samples were speed-vacced to ~50% of the starting volume in order to increase the DNA concentration. DNA concentration was determined by the Invitrogen Qubit fluorometer (Carlsbad, CA).

2.3 DNA Extraction from Soil

Soil was collected in the downtown areas of both Oakland, CA and San Francisco, CA. Four samples were collected in each city at various sites. Samples were extracted using the MoBio UltraClean Soil DNA Isolation Kit #12800. The manufacturer's Alternative Protocol (For Maximum Yields) was followed for this work. The only deviation from the protocol was to wash twice (Step 15) with Solution S4 instead of just once as the protocol stated.

Following extraction, 1ng of each extracted DNA was used in a Real-Time PCR assay to test for inhibition. All samples showed a high level of inhibition of PCR. Based on this each extracted DNA was re-extracted starting from Step 12 of the MoBio Alternative Protocol. This additional extraction is intended to remove additional humic acid. DNA concentration was determined by the Invitrogen Qubit fluorometer.

2.4 DNA Extraction from EPA gauze wipes

Gauze wipes were obtained from a group at LLNL that conducted tests with the EPA. Wipes were used to wipe dirty surfaces indoors and were then inoculated with 6.3×10^7 CFU of *B. thuringiensis* kurstaki spores. Following inoculation, DNA was extracted from the wipes using the Promega Blood Extraction Kit (Madison, WI) according to manufacturer's instructions. After extraction, 5.0×10^4 CFU were amplified and labeled following the procedure in Section 3 below.

2.5 *Bacillus anthracis* Ames DNA spiked in environmental samples

B. anthracis Ames DNA was acquired from the select agent laboratory within LLNL. Sterility test was performed to ensure the DNA is sterile before the DNA was transferred to our laboratory. DNA was quantified using the Invitrogen Qubit fluorometer and copy number was determined. Six concentrations of *B. anthracis* Ames DNA were made in 10 fold serial dilutions from 1-100,000 copies. Each concentration was mixed with 100 pg of extracted DNA from the Spring NCR filters for the aerosol spike experiments or 1 ng extracted DNA from the combination of soil from Oakland and San Francisco for the soil spike experiments.

3. Microarray Processing

An overview of the microarray process is shown in Figure 1.

3.1 Whole Genome Amplification and Purification

The environmental *B. anthracis* Ames spiked samples were amplified using the Qiagen REPLI-g Midi Kit #150043 (Valencia, CA). This kit is intended to provide uniform whole genome amplification using Multiple Displacement Amplification. Each copy number dilution of *B. anthracis* DNA spiked in either 1ng of soil or 100 pg of aerosol DNA was amplified using this kit according to manufacturer's instructions. Samples were allowed to amplify for 16 hours at 30 °C. Amplified samples were purified using the Qiagen Qiaquick PCR Purification Columns #28106 according to manufacturer's instructions. Samples were eluted in 40µL of Buffer EB from the Qiagen kit.

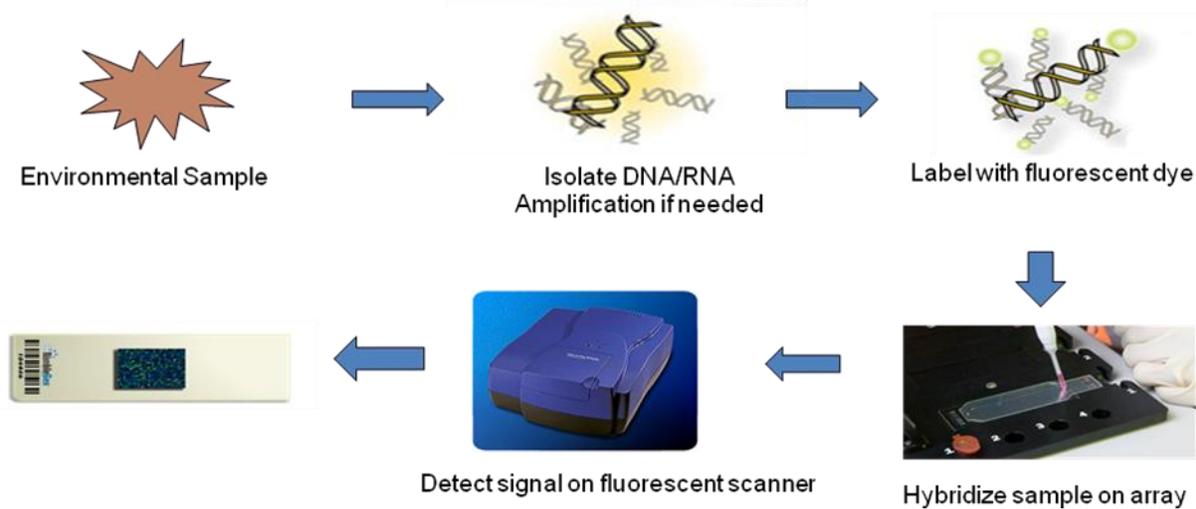


Figure 1. Microarray hybridization process

3.2 Microarray hybridization

The entire 40µL of amplified product was fluorescently labeled using the Roche NimbleGen One-Color DNA Labeling Kit #0522355001 (Madison, WI) according to the recommended protocols. The DNA was purified after labeling, and hybridized using the NimbleGen Hybridization Kit (Cat. 05583683001) to the LLNL Virulence Array according to manufacturers' instructions. The microarrays were allowed to hybridize for 17 hours and washed using the NimbleGen Wash Buffer Kit #05584507001 according to manufacturer's instructions. Microarrays were scanned on an Axon GenePix 4000B 5 µm scanner from Molecular Devices (Sunnyvale, CA). The scanned tif image files were aligned using the NimbleScan Version 2.4 software and pair text files were exported for data analysis.

4. Microarray data Analysis

A maximum likelihood analysis method was used to analyze the microbial hits from samples hybridized to the array. The method was recently published in (3). An example of the analysis results is shown in Figure 2 where *B. thuringiensis* israelensis was run on the Virulence Array. The right-hand column of bar graphs shows the unconditional and conditional log-odds ratios for

each target genome listed at right. The unconditional log-odds is the larger of the two scores; thus the lighter and darker-colored portions represent the unconditional and conditional scores respectively. Targets are color-coded and grouped by taxonomic family, according to the legend at bottom; they are listed within families in decreasing order of conditional log-odds ratio scores. Targets predicted as likely to be present are indicated in red text. The vertical orange dashed line marks 0 on the log-odds ratio scale.

The left-hand column of bar graphs shows the expectation (mean) values of the numbers of probes expected to be present given the presence of the corresponding target genome. The larger “expected” score is obtained by summing the conditional detection probabilities for all probes; the smaller “detected” score is derived by limiting this sum to probes that were actually detected. Because probes often cross-hybridize to multiple related genome sequences, the numbers of “expected” and “detected” probes often greatly exceed the number of probes that were actually designed for a given target organism.

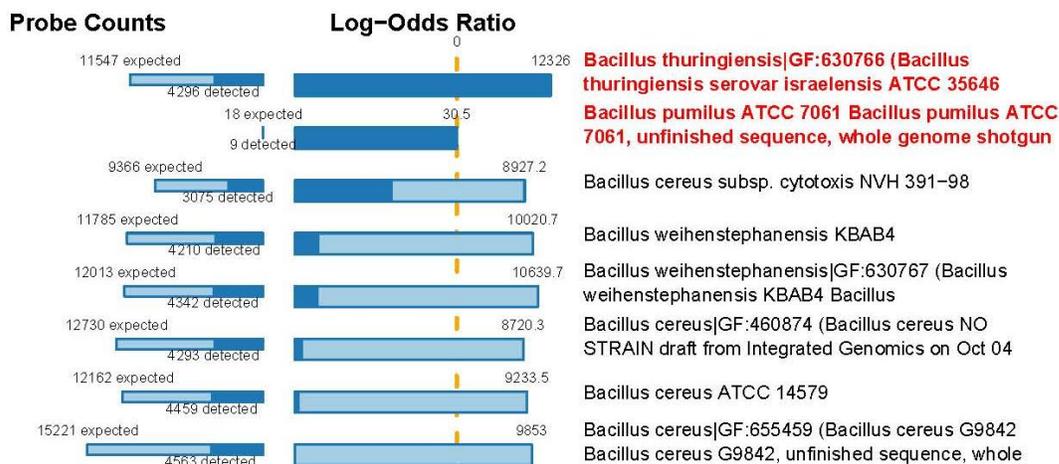


Figure 2. Virulence Chip results for *B. thuringiensis kurstaki*

The probe count bar graphs are designed to provide some additional guidance for interpreting the prediction results. For example, we see that the prediction for presence of *B. pumilus* ATCC 7061 is based on only 9 detected probes; it is therefore given much less weight than the prediction for *B. thuringiensis serovar israelensis*, which is based on over 4000 probes

Results

1. Hybridization of *B. anthracis* and *B. thuringiensis* DNAs on the Virulence Array.

We isolated genomic DNAs from *B. anthracis* Ames, Sterne and A0382 and DNAs from *B. thuringiensis* Israrensis HD500, kurstaki ATCC 33679. The DNAs were hybridized to the Virulence Array. The data is shown in Table 3. The Virulence Array corrected identified each of the DNA to at least the species level. For *B. thuringiensis*, the identification was at the strain level. The probes for this array were designed for species level identification based the detection of probes designed on known virulence genes and antibiotic resistance genes. So it is not unexpected that the *B. anthracis* strains were not correctly detected to the strain level.

Table 3. Virulence Array results from *B. anthracis* and *B. thuringiensis* DNAs

DNA hybridized on array	Virulence Array top hit
<i>B. anthracis</i> Ames	<i>B. anthracis</i> USA6153
<i>B. anthracis</i> Sterne	<i>B. anthracis</i> A0174
<i>B. anthracis</i> A0382	<i>B. anthracis</i> USA6153
<i>B. thuringiensis</i> Israelensis HD500	<i>B. cereus</i> G9842 <i>B. thuringiensis</i> israelensis plasmid pBtoxis
<i>B. thuringiensis</i> Kurstaki ATCC 33679	<i>B. cereus</i> ATCC14579 <i>B. thuringiensis</i> kurstaki <i>B. thuringiensis</i> plasmid pBMB67

2. Determination of the limit of detection of the Virulence Array using *B. anthracis* Ames spiked into BioWatch aerosol samples.

We performed limit of detection testing of the Virulence Array using serially diluted *B. anthracis* Ames spiked into BioWatch aerosol filter extracts that have been subjected to whole genome amplification. Duplicate experiments were run to ensure repeatability and data consistency. 1, 10, 100, 1,000, 10,000, and 100,000 copies of *B. anthracis* Ames were tested. Table 4 below shows results of the virulence array at each of the *B. anthracis* DNA concentration. When 10 copies of *B. anthracis* DNA were spiked into aerosol sample, only one of the two replicate experiments detected *Bacillus cereus*, a very close near neighbor to *B. anthracis*, suggesting that there were not enough probes specific to *B. anthracis* detected at this concentration. This experiment suggested that our detection limit for *B. anthracis* Ames could be in the range between 10-100 copies when the DNA was spiked into 100 pg of aerosol DNA sample.

Table 4. Limit of detection of *B. anthracis* Ames DNA spiked in Biowatch aerosol samples.

Amount aerosol filter DNA	100 pg	100 pg	100 pg	100 pg	100 pg	100 pg
Amount <i>B. anthracis</i> DNA	560 pg	56 pg	5.6 pg	560 fg	56 fg	5.6 fg
<i>B. anthracis</i> DNA Copy #	100,000 copies	10,000 copies	1000 copies	100 copies	10 copies	1 copy
% BA DNA in aerosol DNA	98.2%	35.9%	5.3%	0.56%	0.06%	0.006%
Virulence Array top hit	<i>B. anthracis</i>	<i>B. anthracis</i>	<i>B. anthracis</i>	<i>B. anthracis</i>	<i>B. cereus</i>	Not detected

3. Determination of the limit of detection of the Virulence Array using *B. anthracis* Ames spiked into soil samples.

We performed as similar limit of detection testing of the Virulence Array using serially diluted *B. anthracis* Ames spiked into soil extracts that have been subjected to whole genome amplification. The soils were collected locally in San Francisco and Oakland. Duplicate

experiments were run to ensure repeatability and data consistency. 1, 10, 100, 1,000, 10,000, and 100,000 copies of *B. anthracis* Ames were tested. Table 5 below shows results of the virulence array at each of the *B. anthracis* DNA concentration. When 100 copies of *B. anthracis* DNA were spiked into soil sample, only one of the two replicate experiments detected *Bacillus cereus*, a very close near neighbor to *B. anthracis*, suggesting that there were not enough probes specific to *B. anthracis* detected at this concentration. This experiment suggested that our detection limit for *B. anthracis* Ames could be in the range between 100-1000 copies when the DNA was spiked into 1 ng of soil DNA sample.

Table 5. Limit of detection of *B. anthracis* Ames DNA spiked in soil samples.

Amount soil DNA	1 ng	1 ng	1 ng	1 ng	1 ng	1 ng
Amount <i>B. anthracis</i> DNA	560 pg	56 pg	5.6 pg	560 fg	56 fg	5.6 fg
<i>B. anthracis</i> DNA Copy #	100,000 copies	10,000 copies	1000 copies	100 copies	10 copies	1 copy
% BA DNA in soil DNA	35.9%	5.3%	0.56%	0.06%	0.006%	0.0006%
Virulence Array top hit	<i>B. anthracis</i>	<i>B. anthracis</i>	<i>B. anthracis</i>	<i>B. cereus</i> ***	Not detected	Not detected

4. Analysis of DNA samples from previous DHS or EPA exercises to release and detect *B. thuringiensis* spores.

Aerosol filters were collected during the days when *B. thuringiensis* kurstaki was sprayed to control Gypsy Moth in a DHS surrogate study. Gauge wipes were used to wipe dirty indoor surfaces and inoculated with *B. thuringiensis* kurstaki spores in an exercise conducted with EPA. Genomics DNAs were extracted from filters or wipe samples and run on the Virulence Array. The results are shown in Table 6. We were able to positively identify *B. thuringiensis* kurstaki using the Virulence Array from both the air filter samples and the gauge wipe samples.

Table 6. Detection of *B. thuringiensis* from environmental air or wipe samples

Sample	Air filters collected around Gypsy Moth control study	Gauge wipes collected during an EPA exercise
<i>B. thuringiensis</i> kurstaki specific Taqman assay	Average Ct = 22.77 ± 0.27	Average Ct = 30.70 ± 0.31
Virulence array top hits	<i>B. cereus</i> ATCC 14579 <i>B. thuringiensis</i> kurstaki <i>Burkholderia phymatum</i> STM815 <i>Ralstonia pickettii</i> 12J <i>P. aeruginosa</i> LES	<i>B. cereus</i> ATCC 14579 <i>B. thuringiensis</i> kurstaki <i>Delftia acidovorans</i> <i>S. aureus</i> str. JKD6009

References:

1. Jaing, C., Gardner, S.N., McLoughlin, K., Mulakken, N., Alegria-Hartman, M., Banda, P., Williams, P., Gu, P., Wagner, M., Manohar, C. *et al.* (2008) A functional gene array for detection of bacterial virulence elements. *PLoS ONE*, **3(5)**, e2163. doi:2110.1371/journal.pone.0002163.
2. Allen, J.E., Gardner, S.N. and Slezak, T.R. (2008) DNA signatures for detecting genetic engineering in bacteria *Genome Biology*, **9**, 56.
3. Gardner, S., Jaing, C., McLoughlin, K. and Slezak, T. (2010) A Microbial Detection Array (MDA) for viral and bacterial detection. *BMC Genomics*, **11**, 668doi:610.1186/1471-2164-1111-1668.